



ДЕРЖАВНА СЛУЖБА
ІНТЕЛЕКТУАЛЬНОЇ
ВЛАСНОСТІ
УКРАЇНИ

УКРАЇНА

(19) **UA** (11) **99921** (13) **C2**
(51) МПК
G06F 17/30 (2006.01)

(12) ОПИС ДО ПАТЕНТУ НА ВІНАХІД

(21) Номер заявки: а 2010 01694	(72) Винахідник(и): Панченко Борис Євгенійович (UA)
(22) Дата подання заявки: 17.02.2010	(73) Власник(и): Панченко Борис Євгенійович, вул. Арсенальна, 20, кв. 15, м. Київ, 01011 (UA)
(24) Дата, з якої є чинними права на винахід: 25.10.2012	(56) Перелік документів, взятих до уваги експертизою: Codd E.F. A Relational Model of Data for Large Shared Data Banks. - Comm. ACM, 13, 6 (jun), 1970, p. 377 - 387 Chen P.P. The Entity-Relationship Model: toward a unified view of data. - ACM Trans, on Data base systems, 1:1, 1976, h. 9 - 36 Ontology Change: classification and survey. Flouris Giorgos, Monakenates Dimitris, Kondylakis Haridimos, Plexousakis Dimitris, Antoniou Griroris; Knowl. Eng. Rev., 2008, 23, No 2, с 117-152, Библ. 144 UA, 63036 C2, 15.01.2004 Ronald Fagin. Multivalued Dependencies and a New Normal Form for Relatknal Databases. ACM Transactions on Database Systems, Vol. 2, No. 3, September 1977, Pages 262-278 Ronald Fagin. A Normal Form for Relational Databases That Is Based on Domains and Keys. ACM Transactions on Database Systems, Vol. 6, No. 3, September 1981, Pages 387-415 US, 7092870 B1, Aug.15, 2006 US, 5319745 A, Jun.7, 1994 US, 20040010410 A1, Jan.15, 2004 Матеріали та програми IV міжвузівської науково-технічної конференції. Суми. - Видавництво СумДУ, 2010. - С.37-42
(41) Публікація відомостей про заяву: 25.08.2011, Бюл.№ 16	
(46) Публікація відомостей про видачу патенту: 25.10.2012, Бюл.№ 20	

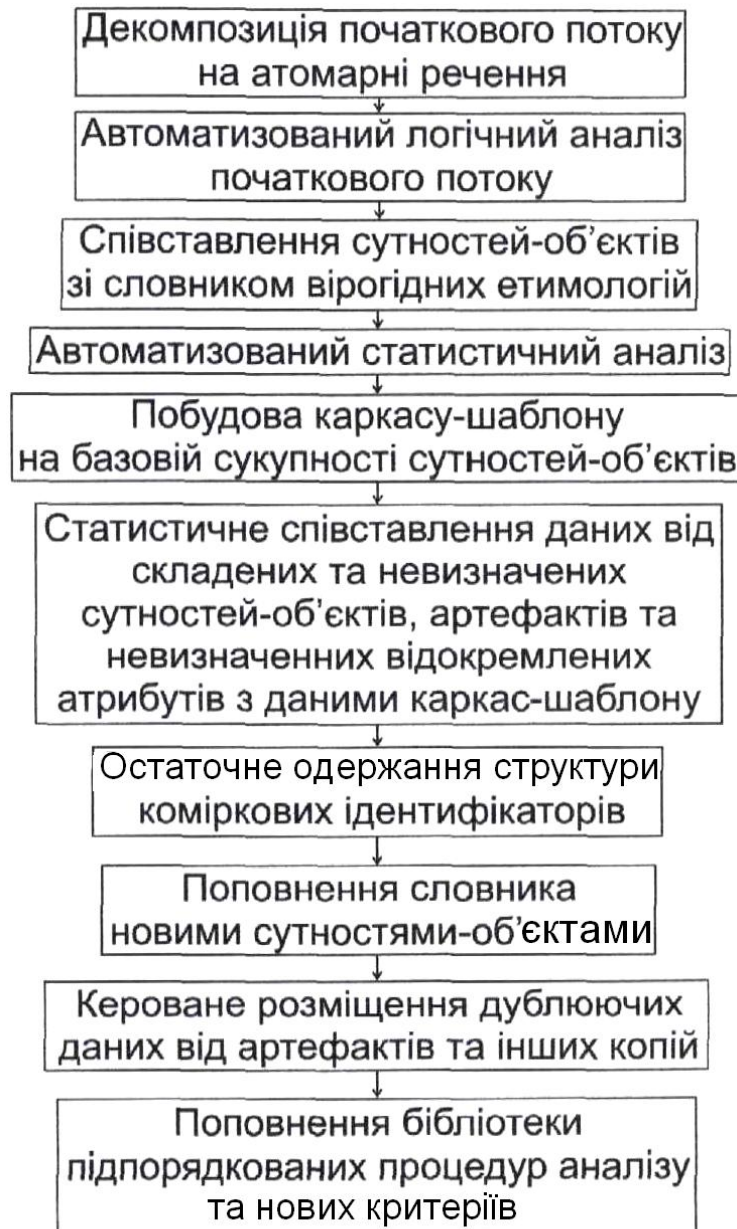
(54) СПОСІБ ПОПЕРЕДНЬОЇ КАРКАСНОЇ СЕПАРАЦІЇ ДАНИХ ПЕРЕД ЇХ МОДИФІКАЦІЙНО ЗДАТНИМ РОЗМІЩЕННЯМ У СХОВИЩЕ АБО ПРОЦЕСОМ ПОДАЛЬШОЇ ОБРОБКИ

(57) Реферат:

Спосіб попередньої каркасної сепарації даних перед їх модифікаційно здатним розміщенням у сховище або процесом подальшої обробки належить до інформаційних технологій і може бути використаним для побудови пристроїв розпізнавання мови, пристроїв-перекладачів, експертних систем тощо. У способі всю сукупність сутностей-об'єктів, що відносяться до певної предметної області, автоматизовано розподіляють на п'ять груп: атомарних, складених та слабких сутностей-об'єктів, а також артефактів, тобто сутностей-копій, дані від яких умовно розміщуються у сховище, та групу невизначених сутностей-об'єктів, семантика яких підлягає подальшому уточненню. Спосіб передбачає можливість поповнення групи алгоритмів сепарації,

UA 99921 C2

кожний з яких дає можливість більш точного віднесення тієї чи іншої сутності-об'єкта до згаданих груп. А використання їх послідовно дає можливість максимально зменшити п'яту групу - групу невизначених сутностей-об'єктів, які одночасно мають суперечливі ознаки. У способі заявляється декілька алгоритмів. Це алгоритм, заснований на використанні словника сутностей-об'єктів, який розміщено в загальноновживаних мережах та постійно поповнюється, на функціональних залежностях між даними від сутностей-об'єктів, що дозволяють зіставити сутності-об'єкти між собою, алгоритм відстежування повторних сутностей-об'єктів в бінарних парах, алгоритм статистичного аналізу детермінованих або багатозначних залежностей, а також алгоритм послідовних наближень та модифікацій на каркасі-шаблоні зв'язків. Технічним результатом є можливість одночасно використовувати реляційну та об'єктно-орієнтовану моделі розміщення даних таким чином, що кожна сутність з реляційної моделі є тотожною об'єкту та сприяння вирішенню питання стандартизації побудови сховищ даних.



Фіг. 1

Винахід належить до галузі інформаційних технологій і може бути використаним для побудови пристроїв розпізнавання мови, пристроїв-перекладачів, експертних систем, систем автоматизованого аудиту коректності роботи впроваджених інформаційних комплексів, а також систем автоматизованого проектування сховищ для зберігання та обробки даних з можливістю

5 гнучкої модифікації структури сховища для довільної предметної області, тобто предметної області довільного обсягу та довільної структури (надалі - просто "предметної області").

Тут під терміном "дане" розуміється матеріальний електричний заряд певної величини або матеріальне електромагнітне поле певної потужності, а під тим чи іншим способом маніпулювання даними розуміється такий керований матеріальний вплив на відповідне

10 матеріальне середовище - наприклад, інше електромагнітне поле, яке в свою чергу керує даними, що призводить до певного їх розміщення в цифровій пам'яті - тобто матеріальному ж таки середовищу, яке може бути побудовано згідно з типовими відомими принципами як сукупність конденсаторів, тригерів, магнітних шарів й т. і. Тому той факт, що маніпулювання даними - це вплив "матеріального на матеріальне", в міжнародному патентному класифікаторі

15 підтверджено виділенням цього процесу окремого класу G06F.

Широко відомі традиційні способи розміщення даних, які базуються на класичних технологіях [Codd E.F. A Relational Model of Data for Large Shared Data Banks. - Comm. ACM, 13, 6 (jun), 1970, p. 377-387; Codd E.F. Normalised Data Base Structure: a Brief Tutorial. - Proc. ACM, SIGFIDET, 1971, Workshop, San Diego, Calif, Nov. 1971, p. 1-18, Maier D. Why isn't there an object-oriented data model? - Proceedings IFIP 11th World Computer Conference, San Francisco, C A, August-September, 1989, Chen P.P. The Entity-Relationship Model: toward a unified view of data. - ACM Trans, on Data base systems, 1:1, 1976, h. 9-36]. Ці методи мають основний недолік - вони не вирішують питання отримання універсальної та гнучкої структури сховища, а також створюють сховище залежним від початкової семантики предметної області і не вирішують

20 питання гнучкої модифікаційної здатності у подальшій експлуатації. Що ж до використання методики онтологій, тобто побудови параметризованих тлумачних словників предметних областей, суттєвий огляд способів та підходів викладено в публікації "Зміни онтологій: класифікація та огляд" [Ontology Chenge: classification and survey. Flouris Giorgos, Monakenates Dimitris, Kondylakis Haridimos, Plexousakis Dimitris, Antoniou Griroris; Knowl. Eng. Rev., 2008, 23, No 2, С. 117-152, Библ. 144]. Проте в усіх цих підходах не розглядається питання побудови способу, який дозволяє автоматизовано створювати модифікаційно здатні гнучкі сховища на базі звукового або текстового опису предметної області природною мовою.

Близьким до способу, що заявляється, є спосіб використання попереднього формального опису предметних областей, що відтворений в широковідомій онтології WordNet [Соловьев В.Д., Добров Б.В., Иванов В.В., Лукашевич Н.В., Онтологии и тезаурусы, МГУ, Москва, 2006]. Проте ця онтологія також має суттєвий недолік - в ній відсутній єдиний фактор, що дозволяє систематизувати семантику сутностей-об'єктів, тобто іменників з опису предметної області, а також відсутній підхід, що доказово мінімізує кількість базових категорій, які дозволяють вести

35 автоматизовану сепарацію сутностей-об'єктів зі значної кількості термінів та синонімів з початкового потоку опису предметної області.

Проте, хоча всі наведені системи мають означені недоліки, своїм існуванням вони доводять можливість реалізувати винахід, що заявляється. Ці відомі продукти та впроваджені в згаданих галузях інструментальні засоби суттєво відрізняються принципами побудови та підходами до маніпулювання даними як один від одного, так і від способу, що заявляється. Але ці значні

45 відмінності ніяк не зменшують можливість реалізувати спосіб та ніяк не впливають на мету винаходу.

Найбільш близьким до запропонованого є спосіб розміщення даних у цифровому сховищі, яке побудовано у відповідності до декартового перемноженням сурогатних ключів сутностей-об'єктів [Панченко Б.С. Спосіб розміщення даних у комп'ютерному сховищі із забезпеченням модифікаційності його структури. Патент України №63036 від 15.01.2004]. За цією моделлю

50 завдяки декартовому перемноженню множин сурогатних ключів сутностей-об'єктів формують сховище із системи реляційних таблиць, які заповнюють даними - атрибутами сутностей-об'єктів та атрибутами зв'язків. Проте цей спосіб має певний недолік - він не дозволяє автоматизовано відокремити з початкового потоку опису предметної області різноманітні замасковані семантики сутностей-об'єктів.

55

В матеріалах цієї заявки всі терміни та поняття, що не є загальноновживаними, зведено в окремий тлумачний словник та розміщено в завершальній частині опису.

Отже, в способі, що заявляється, всі сутності-об'єкти розподіляють на п'ять категорій. Перша - це атомарні сутності-об'єкти, які в деяких моделях даних іноді називають базовими. Друга

60 категорія - слабкі сутності-об'єкти, які функціонально залежать від атомарних і в моделях даних

мають аналогічну назву. При цьому така залежність може бути як лише на рівні ідентифікації слабких атрибутів, так і на рівні всього існування залежних слабких сутностей-об'єктів. Проте існує виключення. Для певних предметних областей деякі слабкі сутності-об'єкти можуть бути примусово призначені як атомарні. При цьому користувач призначає цю сутність-об'єкт як

5

останню ланку в її ієрархії. І їй штучно призначається ідентифікатор, який унікально ідентифікує всі атрибути. Такі виключення є своєрідним обмежуючим колом предметної області, коли користувачеві відомо, що протягом значного часу експлуатації сховища даних, яке проектується або обстежується користувачем, це коло не буде розширюватись. Проте саме такі виключення і призводять до неможливості здійснювати модифікації системи експлуатації сховища.

Третя категорія - складені післязв'язкові сутності-об'єкти, які в моделях даних мають ще й назву багатобічних. Отже, в цьому способі сутності-об'єкти утворюються за наступною схемою: на базі атомарних породжуються слабкі, тобто функціонально залежні від базових. А на сукупності атомарних та слабких сутностей-об'єктів завдяки утворенню різноманітних зв'язків між ними створюються складені післязв'язкові сутності-об'єкти. При цьому означений процес

10

15

20

утворення слабких та складених сутностей-об'єктів маскують частини мови - іменники, віддієслівні іменники, різноманітні терміни, що їм відповідають, категорій, що їх узагальнюють й т.і. Саме це і робить автоматизовану сепарацію актуальною. Адже переважна більшість складених сутностей-об'єктів як правило помилково потрапляє до категорії слабких, або навіть атомарних, що, в свою чергу, знову ж таки призводить до збільшеної жорсткості системи та неможливості її гнучкого розвитку без докорінних переробок.

До четвертої категорії спосіб відносить артефакти, тобто сутності-копії, дані від яких будуть умовно розміщуватися у сховище згідно з рішенням користувача. До артефактів, наприклад, може бути віднесений будь-який документ, який користувач предметної області створює саме заради того, щоб скопіювати ті чи інші атрибути тих чи інших сутностей-об'єктів. При цьому не

25

просто скопіювати атрибути однієї конкретної сутності-об'єкта, а ще й об'єднати в цій новій штучно створеній сутності-об'єкті декілька атрибутів від різних сутностей-об'єктів.

Артефакти - це як правило наслідкові сутності-об'єкти. Тому реєструючи їх в системі, що експлуатує сховище, користувач наражається на значне дублювання даних. А це, в свою чергу, призводить до потреби додаткового відстежування цілісності ще й надлишкових даних. Виключення становить сукупність штучних сутностей-об'єктів, кожна з яких об'єднує лише певну частину атрибутів іншої, більш загальної не штучної сутності-об'єкта. При цьому об'єднання сукупностей атрибутів кожної штучної сутності-об'єкта строго тотожне сукупності всіх атрибутів загальної, нештучної сутності-об'єкта. Тобто, жодна зі штучних сутностей-об'єктів не має жодного атрибута, що є спільним хоча б для двох штучних сутностей-об'єктів. А також не існує жодного атрибута у загальної нештучної сутності-об'єкта, для якого не існувало б копії серед сукупності штучних. Тож, цю сукупність штучних сутностей-об'єктів спосіб також віднесе до категорії "артефакти". Проте від стеження цілісності таких дубльованих даних спрощується.

30

35

І завершує перелік група невизначених сутностей-об'єктів, семантика яких підлягає подальшому уточненню.

40

Прикладом атомарної може бути сутність-об'єкт "людина", "всесвіт", "собака", "кішка" і т.і. При цьому приналежність цих сутностей-об'єктів до певних подальших категорій - так звана класифікація атомарних сутностей - є штучною семантичною надбудовою користувача, яка і й маскує зміст сутності-об'єкта. Прикладом слабкої можуть бути сутності-об'єкти "підрозділ", "відділ", "лабораторія", "квартира", адже кожна з цих сутностей-об'єктів не самодостатня. І в довільній предметній області функціонально залежить від більш старших - предкових - сутностей-об'єктів. Прикладом складеної сутності є будь-яка подієва сутність-об'єкт - "екзамен", "концерт", "виставка", "угода", "мітинг" й т.і., адже її зміст являє собою "продукт" рівноправної взаємодії декількох інших сутностей-об'єктів. Прикладом артефактів може бути "накладна", "рахунок" (до сплати в ресторані, за інші послуги), "акт" й т.і.

45

50

Спосіб, що заявляється, побудовано у відповідності до теорії каркасної моделі предметної області [Панченко Б.Е., О синтезе универсальной логической модели данных // Вестник СумГУ, - Сумы. - 2009. - Серия "Техн.". Вып. 2. - С. 60-66 та Панченко Б.Е., Писанко И.Н., Свойства реляционного каркаса на множестве семантически атомарных предикатов.// Кибернетика и системный анализ. - К.: 2009. - № 6. - С. 120-129]. В цій моделі основним інструментом аналізу предметної області є багатомісні семантично атомарні предикати, що базуються на єдиному факторі - походженні сутності-об'єкта. При цьому походженні не терміну, а саме змісту, який кодується цим терміном. В цій моделі використовується той факт, що для предметної області завжди існує обмежена базова сукупність сутностей-об'єктів, до якої належать лише атомарні та слабкі сутності-об'єкти. А всі інші сутності-об'єкти (яких найчастіше - набагато більше) синтезуються на цій сукупності завдяки каркасу зв'язків, тобто булеану всіх підмножин зв'язків

55

60

сутностей-об'єктів з базової сукупності. Тобто, решта сутностей-об'єктів є наслідком функціонування цієї предметної області.

Отже, в загальному вигляді алгоритм способу зводиться до наступних кроків.

1. Автоматизоване вилучення базової сукупності сутностей-об'єктів, яка в початковому потоці опису предметної області може бути замаскованою різноманітними термінами, категоріями, допоміжними іменниками, синонімами й т. і. Базова сукупність відокремлюється від артефактів, невизначених та складених сутностей-об'єктів. При цьому це здійснюється методом послідовних наближень, коли на кожному наступному кроці завдяки певним логічним та математичним критеріям здійснюється уточнення кожної попередньої сукупності даних. Для цього спосіб передбачає послідовне або паралельне виконання для кожної сутності-об'єкта процедури автоматизованого логічного порівняння з кожною іншою сутністю-об'єктом. А кількість підлеглих логічних критеріїв для порівнянь нічим не обмежена і може бути відокремлена у зовнішню бібліотеку, що поповнюється.

2. Синтез каркасних еталонних складених сутностей-об'єктів - побудова на базовій сукупності за допомогою булеану зв'язків за принципом "всі з усіма" каркасу-шаблону.

3. Остаточна сепарація складених сутностей-об'єктів завдяки процедурам статистичного порівняння отриманих на каркасі-шаблоні еталонних складених сутностей-об'єктів та тих складених сутностей-об'єктів, що на завершальному етапі відокремлені в початковому потоці. Адже саме складені сутності-об'єкти є в предметній області найбільше замаскованими та такими, що мають найбільш суперечливе походження змісту.

4. Рекомендації адміністрації словника імовірних етимологій про можливість поповнення його ресурсів новими групами сутностей-об'єктів, якщо в остаточних групах ніяких суперечностей не виявлено.

Тож при більш детальному розгляді, спосіб попередньої каркасної сепарації даних перед їх модифікаційно здатним розміщенням у сховище або процесом подальшої обробки полягає в тому, що розміщувані дані автоматизовано розподіляються на вищезазначені п'ять груп згідно з результатами автоматизованого логічного та статистичного аналізу голосового, текстового або схемного опису певної предметної області. В предметній області виявляються сутності-об'єкти, які об'єднують кожну таку групу. І така група даних має спільний набір характеристик, які відповідають спільному предикату. А групи сутностей-об'єктів перебувають між собою лише або у рівноправних, або в ієрархічних відносинах.

Спосіб передбачає, що опис предметної області, яка підлягає такому автоматизованому датологічному моделюванню, має бути вираженням наступною мовною формою: одиницею зчитування є атомарне речення (надалі - просто "речення"), що містить пару сутностей-об'єктів, що кодуються іменниками з унікальним політерним написом. При цьому передбачається, що іменники, що повторюються, означають одну й ту саму сутність-об'єкт. Тому таке повторення в межах одного речення означатиме тривіальну пару, тобто таку, що несе лише інформацію про існування сутності-об'єкта в предметній області без зв'язків її з іншими. І декларує її для подальших кроків аналізу.

А дієслово з унікальним політерним написом символізує виключно бінарний зв'язок між ними, тобто зв'язок між парою сутностей-об'єктів цього ж речення. Передбачається, що дієслова, що повторюються в різних реченнях, означають один і той самий клас зв'язку. Тому основна місія атомарного речення - інформувати про наявність сутностей-об'єктів в певній предметній області та декларувати клас зв'язку цієї пари. Речення, що містять в собі більше, ніж пару сутностей-об'єктів, є складеними. Вони підлягають автоматизованій декомпозиції. Для цього може використовуватися будь-який відомий алгоритм декомпозиції складених речень. Наприклад той, що використовується в будь-якому компіляторі як алгоритм розбору рядків. Проте, ті складені речення, які неможливо автоматизовано декомпонувати до бінарної форми за технологічних причин - наприклад, відсутності чіткої структури, що об'єднує їх в одне складене речення, - з початкового потоку опису вилучаються та відокремлюються до фрагменту опису, що підлягає подальшому уточненню.

Спосіб не передбачає верхнього обмеження кількості речень. А нижня кількість обмежується змістом предметної області. Проте передбачається попередній формальний аналіз наявності для кожної задекларованої сутності-об'єкта хоча б одного зв'язку з будь якою іншою сутністю-об'єктом.

Отже, на першому кроці способу здійснюється зчитування звукового голосового сигналу в реальному часі або файла з записаним голосовим сигналом, що надиктовані природною мовою та описують предметну область. Опис може бути підготовленим у вигляді текстового файлу, сформованого текстом природною мовою, або у вигляді файла, сформованого мовою послідовних схем чи графів, що відповідають опису предметної області. Це також може бути

послідовність файлів сховищ даних, що вже існують та впроваджені в експлуатацію, задля дослідження вірогідних протиріч в схемах даних та прогнозування витрат на модифікації при подальшому розвитку впровадженої системи. При цьому для перетворення файла вхідного датового опису предметної області, сформованого мовою послідовних схем чи графів, у потік слів, спосіб вимагає кожній графовій фігурі схеми - наприклад, прямокутнику, ставити у відповідність іменник, а дузі графу, позначену на схемі прямою або кривою лінією, що з'єднує ці прямокутники, ставити у відповідність дієслово. Спосіб передбачається окрема процедура строгого відокремлення зі схемного початкового потоку пар сутностей-об'єктів та їх зв'язків, а також позначення їх іменниками та дієсловами, тобто опрацювання графових схем типу ER-схем з урахуванням обмежень унікальності політерного найменування сутностей-об'єктів. Аналогічна процедура використовується і при перетворенні файлів сховищ даних, що вже експлуатуються. Ці види файлів також зчитуються.

Для подальшого аналізу кожний потік може використовуватись не лише відокремлено, а й у відповідності один до іншого. Після цього завдяки відомим процедурам здійснюється розпізнавання відокремлених слів у звуковому потоці або перетворення в словесний потік сукупності схем чи файлових структур сховищ даних, а після цього - розміщення всіх одержаних слів в пам'яті.

На подальшому кроці почергово аналізується кожне слово за принципом послідовних наближень, причому існує можливість втручання користувача завдяки роботі способу в діалоговому режимі, що дає можливість динамічно враховувати додаткові відомості про дані з предметної області. Неструктурований сукупний початковий потік, що описує предметну область від користувача, в пам'яті перетворюється у потік, що має вищезгадану спеціалізовану датову форму та структуру, в якій технологічною одиницею аналізу є одне атомарне речення.

Для подальшої реалізації способу в пам'яті формується ділянка, де розміщуються структуровані коміркові ідентифікатори, структура кожного з яких не довільна, не задана користувачем та не отримана якимось іншим відмінним способом, а строго відповідає імовірній семантичній структурі змісту кожної сутності-об'єкта. Ця структура відповідає структурі предиката, що утворює сутність-об'єкт. Для автоматизованого вищелювання замаскованої структури використовуються логічні та математичні критерії, які побудовано у відповідності до закономірностей, виявлених в предметних областях каркасної моделі даних. В підґрунті цих критеріїв - єдиний узагальнений фактор - походження змісту сутності-об'єкта, тобто етимологія її змісту (надалі - просто "етимологія").

Отже в способі, що заявляється, використовується те, що всі інші фактори, що характеризують семантику будь-якої сутності-об'єкта в предметній області є функціонально залежними від етимології. Етимологія, в свою чергу, описується математичною логікою предикатів та у вигляді рядкового структурованого ідентифікатора має наступну загальну схему:

$$X_1^{m_1} + X_2^{m_2} + X_3^{m_3} + \dots + X_{kj}^{m_k},$$

де кожна ланка $X_{kj}^{m_k}$ - відокремлений ідентифікатор факту походження j -ї сутності-об'єкта, k_j - номер ланки ідентифікатора j -ї сутності-об'єкта (підрядковий індекс), m_k - номер відповідної породжуючої сутності-об'єкта з об'єднаної групи атомарних та слабких сутностей-об'єктів (надрядковий індекс), причому кожне m_k може отримати будь яке значення лише з множини $\{1, 2, \dots, N_0, \dots, N\}$, де N_0 - загальна кількість атомарних сутностей-об'єктів, N - сумарна кількість атомарних та слабких сутностей, j - номер довільної сутності-об'єкта в предметній області. При

цьому у випадку повної сукупності зв'язків $j = \{1, 2, \dots, N, \dots, (2^N - 1)\}$. Знак "плюс" в загальному вигляді схеми етимології означає рядкове об'єднання. Для атомарних сутностей етимологією є лише одна ланка X_j^j , в якій $m = j$. Тобто атомарна сутність породжує сама себе. В способі, що заявляється, атомарні сутності-об'єкти одержують в загальній сукупності перші номери, тобто для них $j = 1, N_0$. Для слабких сутностей етимологією є вищезгадана рядкова сума ланок, де

кожному номеру k ланка $X_{kj}^{m_k}$ відповідає строго. Тобто послідовність ланок строго відповідає послідовності залежностей кожної наступної ланки від попередньої, що в свою чергу відповідає послідовності синтезу кожною попередньою слабкою сутністю-об'єктом, аж до найстаршої атомарної, наступної сутності-об'єкта. Для складених сутностей етимологією є вищезгадана рядкова сума ланок, де місце кожної ланки $X_{kj}^{m_k}$ не строге, тобто послідовність ланок не має

значення. Проте сумарна сукупність ланок строго відповідає сукупності утворюючих сутностей-об'єктів. Отже в загальному випадку для будь якої сутності-об'єкта весь структурований комірковий ідентифікатор являє собою сумарний рядок літер або цифр, кожна ланка яких має мінімально достатній рядковий розмір. Отже такий ідентифікатор, наприклад, в реляційній моделі даних може використовуватись як мінімально достатній сурогатний ключ реляційної таблиці, яка об'єднує в одному відношенні всі властивості конкретної сутності-об'єкта. Її атрибути - це аргументи утворюючого багатомісного предиката сутності-об'єкта. При цьому кожне місце в предикаті - це його аргумент, а значить атрибут відповідної сутності-об'єкта, де атрибутів може бути довільна кількість. Це не впливає на структуру функціональної частини предиката, а значить і на структуру коміркового ідентифікатора. Кожна ланка етимології сутності-об'єкта означає зв'язок зі сторонніми сутностями-об'єктами, які брали участь у походженні конкретної сутності-об'єкта, якщо остання являє собою або слабку, або складену, тобто післязв'язкову сутність-об'єкт. Отже, кожна ланка $X_{kj}^{m_k}$ коміркового ідентифікатора будується у строгій відповідності до етимології змісту сутностей-об'єктів з опису предметної області. Кожна сутність-об'єкт в предметній області може відповідати або атомарному, тобто, унарному в функціональній частині, але багатомісному в аргументній частині, предикату, а значить мати унарний ідентифікатор x^j , або складеному в функціональній частині і багатомісному в аргументній частині предикату, тобто мати складений ідентифікатор $X_{kj}^{m_k}$, де сума - по $k_j, k_j = 1, K_j$, отже ідентифікатор має вищезазначену загальну структуру. Складена функціональна частина предиката є наслідком кон'юнкції атомарних предикатів, що і відповідає рядковому об'єднанню множин даних ланок ідентифікаторів, тобто додаванню рядків. При цьому загальна кількість ланок K_j являє собою арність функціональної частини утворюючого багатомісного предиката, яка в загальному випадку може дорівнювати 2, 3, ..., 10 і т. і. А у випадку атомарної сутності-об'єкта дорівнює виключно одиниці.

В подальшому в ідентифіковані комірки сховища можуть розміщувати групи атрибутів сутностей-об'єктів, наприклад, їх найменування та групу інших властивостей, що є аргументами відповідних атомарних або складених багатомісних предикатів. Унарні ідентифікатори комірок сховища строго відповідають атомарним сутностям-об'єктам, а складені ідентифікатори комірок строго відповідають слабким та складеним сутностям-об'єктам.

На подальшому кроці в пам'яті здійснюється послідовне або одночасне, тобто паралельне, виконання для кожної сутності-об'єкта з кожного речення, тобто, з кожної пари, процедури порівняння з кожною іншою сутністю-об'єктом. Ця процедура виконує відокремлені підпорядковані способи автоматизованого логічного вицленювання замаскованої етимології кожної сутності-об'єкта, а значить семантичної структури її змісту. Результат їх виконання - це шукана сепарація, тобто, надання кожній комірці, де зберігаються дані від атрибутів кожної сутності-об'єкта з початкового потоку відповідних структурованих коміркових ідентифікаторів та перегруповання сутностей-об'єктів в сховищі у вищезгадані окремо розміщені групи. При цьому відновлення структури кожної ланки етимології сутностей-об'єктів на цьому кроці здійснюється автоматизованим логічним аналізом іменників та дієслів, тобто аналізом змісту сутностей-об'єктів та змісту зв'язків, без врахування множин конкретних значень конкретних атрибутів сутностей-об'єктів. Аналіз засновано на порівнянні змісту сутностей-об'єктів між собою за принципом "всі з усіма" з використанням словника імовірних етимологій та семантичних структур сутностей-об'єктів, що може бути розміщеним також і в загальноновживаних мережах та постійно уточнюється й поповнюється в автоматичному режимі. В цьому словнику кожному іменнику заздалегідь поставлено у відповідність найімовірнішу структуру функціональної частини предиката, що цей іменник обумовлює. Тобто етимологію його змісту, задану гіпотетично або отриману шляхом сторонніх досліджень і визнану користувачами. При цьому ступінь цієї імовірності залежить від специфіки предметної області. Отже на цьому кроці встановлюється відповідність між словами з початкових потоків та словами, що існують в словнику. Результатом такого співставлення є перше наближення шуканої сепарації сутностей-об'єктів. А також отримання першого наближення структур їх етимологій. При цьому ті слова, що позначають невідомі поки що для словника сутності-об'єкти та класи зв'язків, для подальшого автоматизованого аналізу відокремлюються. А у випадку, якщо невідомих словнику сутностей-об'єктів та зв'язків у початкових потоках не виявлено, автоматизований логічний аналіз завершується.

Всі подальші кроки способу, що заявляється, завдяки різним критеріям відстежують етимологію невідомих словникові сутностей-об'єктів. А також повертають користувачеві певні

рекомендації щодо знайдених логічних помилок та протиріч у початковому потоці щодо можливих некоректних використань іменників та дієслів, що може означати навіть нелогічності в роботі певних ділянок предметних областей. Тому при виявленні розбіжностей користувачеві надаються відповідні висновки.

На наступному кроці проводиться автоматизований логічний аналіз тих сутностей-об'єктів та зв'язків, що виявилися невідомими словнику імовірних етимологій. При цьому передусім відокремлюються невідомі потенційні складені сутності-об'єкти завдяки автоматизованому логічному порівнянню кожної з невідомих сутностей-об'єктів з тими, що утворюються з повторюваних іменників і повторюваних дієслів з початкових потоків завдяки об'єднанню їх в одну складену, тобто багатобічну післязв'язкову сутність-об'єкт. Таке об'єднання можливе за умови збігу класу зв'язку, тобто збігу дієслів між різними парами, тому що саме завдяки багаторазовій повторюваності згаданих іменників в декількох різних зв'язках з одного класу, тобто для декількох однакових дієслів, імовірність того, що ці сутності-об'єкти належать саме до групи складених сутностей-об'єктів значно підвищується. Якщо ж з'ясується, що таке наближення - помилкове, це не внесе значної некоректності. Адже на наступних кроках воно буде уточнюватись. Наявність в цих попередньо сепарованих групах невизначених сутностей-об'єктів, що мають логічні суперечності, та артефактів на цьому кроці способу ігнорується.

На наступному кроці завершується автоматизований логічний аналіз початкового потоку. Останнє логічне порівняння - аналіз групи тих сутностей-об'єктів та зв'язків, що виявилися невідомими словнику імовірних етимологій та залишилися після вилучення потенційно складених сутностей-об'єктів. Від сутностей-об'єктів, що залишилися, відокремлюються невідомі атомарні сутності-об'єкти з використанням єдиного логічного критерію, що в загальному випадку для ідентифікації будь якого значення природного, тобто не штучно призначеного користувачами, атрибута атомарної сутності-об'єкта достатньо лише назви сутності-об'єкта та назви цього атрибута. Це неможливо у випадку слабкої сутності, адже саме слабкість і полягає в тому, що неможливо ідентифікувати будь яке значення будь-якого природного атрибута слабкої сутності-об'єкта без врахування її зв'язку з функціонально залежною, тобто ієрархічно старшою сутністю-об'єктом. На завершальному кроці автоматизованого логічного аналізу кожна сутність-об'єкт, що лишилася від попередніх кроків, отримує статус або атомарної сутності-об'єкта, або слабкої, або невизначеної, причому наявність артефактів на цьому кроці ігнорується і вони також одержують один зі згаданих статусів.

Якщо після автоматизованого логічного аналізу початкового потоку сутностей-об'єктів та зв'язків група невизначених сутностей-об'єктів, які мають суперечливу семантику, не стає порожньою, тобто завдяки автоматизованому логічному аналізу неможливо віднести ці сутності-об'єкти до згаданих трьох категорій, кожній з цих суперечливих сутностей-об'єктів примусово призначають статус атомарної. Але на рівні їхнього коміркового ідентифікатора обов'язково позначають це, додаючи до унарного ідентифікатора спеціалізовану відокремлену ланку, відповідальну за цю особливість. Тим самим в групі атомарних сутностей-об'єктів формується окрема підгрупа суперечливих сутностей-об'єктів, що при подальшій експлуатації сховища при потребі модифікації його структури дає можливість користувачеві внести відповідні корегування.

Тож для подальших кроків спосіб потребує внесення додаткової інформації, якщо вона не була внесена в початкових потоках, щодо не менше двох природних атрибутів кожної з сутностей-об'єктів, що аналізуються. А також декількох (з практики відомо, що як правило - не більше трьох) значень кожного з цих атрибутів.

На наступному кроці від попередньо відібраних груп сутностей-об'єктів остаточно відокремлюються артефакти, тобто сутності-копії. Для цього здійснюється автоматизоване статистичне порівняння, засноване на використанні відомих процедур статистичного аналізу для виявлення детермінованих функціональних або кореляційних чи регресивних багатозначних залежностей між значеннями даних в атрибутах сутностей-об'єктів. Наявність або відсутність таких залежностей дозволяє підтвердити або спростувати прямі збіги груп атрибутів, а також замасковану етимологію та семантичну структуру, одержану на попередніх кроках.

Як свідчать певні дослідження, для відстеження наявності, наприклад, прямих збігів атрибутів-копій достатньо порівняти не більше десяти груп значень, тобто, не більше десяти груп кортежів для реляційного формату зберігання значень атрибутів сутностей-об'єктів. Для відстеження закономірності на цьому кроці способу від кожної сутності-об'єкту достатньо не більше двох природних атрибутів. А для відстеження, наприклад, багатозначної залежності, що спостерігається лише між атрибутами складених сутностей-об'єктів та відокремлено атрибутами кожного з їх предків, що брали участь в утворюючих зв'язках цих післязв'язкових складених сутностей-об'єктів, достатньо порівняти не більше двохсот груп значень. Тобто, не

більше двохсот груп кортежів для реляційного формату зберігання значень атрибутів сутностей-об'єктів. При цьому між кожним сумарним значенням примірників загальної сукупності всіх відокремлених атрибутів предків та значеннями примірників будь якого або навіть кожного з атрибутів складених сутностей-об'єктів виникає вже не багатозначний, а детермінований функціональний зв'язок, якщо саме ці предки утворювали саме цю складену сутність-об'єкт. Наявність такого детермінованого зв'язку є достатнім критерієм для ідентифікації та відокремлення складених сутностей-об'єктів. При цьому для відстеження цієї закономірності від кожної сутності-об'єкта також достатньо не більше двох природних атрибутів.

Проте, для коректності статистичного аналізу вся сукупність значень всіх атрибутів від всіх сутностей-об'єктів предметної області мають відповідати єдиному моменту часу життя предметної області. При цьому відстань між сусідніми проміжками часу має бути достатньою для виникнення дійсно нового стану предметної області, адже якщо ця умова не виконується, закономірності можуть виявитися некоректними.

Отже при виникненні на цьому кроці прямих збігів назв груп атрибутів, а також і збігів їх значень у різних сутностей-об'єктів, спосіб відокремить артефакти. Та на рівні їх коміркових ідентифікаторів окремо зафіксує цей факт, що дасть можливість користувачеві визначитися щодо зберігання надлишкових даних. Проте ситуація, коли назви атрибутів, що належать різним сутностям-об'єктам, є різними, а їх значення за якихось причин ідентичні, з'ясовується також на збільшеній кількості значень атрибутів. Коли їх не менше ста, тоді збіг не може бути випадковим, що знаходить відображення в структурі коміркового ідентифікатору.

На наступному кроці будують уточнене наближення сепарації, для чого відокремлюють групи значень атрибутів, що залежать від часу, та групи значень атрибутів, що від часу не залежать. Або, якщо і залежить, то лише від дуже значних проміжків часу - розвитком та змінами яких можна знехтувати у порівнянні з іншими групами значень атрибутів. При цьому групу атрибутів, яка практично не залежить від часу, відокремлюють до групи сутностей-об'єктів, що створюють структуру предметної області. Адже структура будь якої системи значно повільніше залежить від часу, ніж саме її функціонування, тобто утворювання певних зв'язків між сутностями-об'єктами. Тож на цьому кроці за уточнене чергове наближення складених сутностей-об'єктів береться група сутностей-об'єктів, які залежать від часу. А інша група отримує статус сукупності атомарних, атомарних-невизначених та слабких, адже артефактів початковий потік позбувся на попередніх кроках. І це знаходить відображення у відповідних коміркових ідентифікаторах. Після цього кожна складена сутність-об'єкт зі щойно отриманої групи зіставляється з групою складених сутностей-об'єктів, що залишилися після автоматизованого логічного аналізу. При цьому якщо спостерігаються збіги, коміркові ідентифікатори залишаються без змін. В іншому ж випадку у кожній з потенційно складених сутностей, отриманих різними кроками, утворюються декілька відповідних незалежних коміркових ідентифікаторів, тобто декілька потенційних етимологій, що фіксують цю обставину. А ці сутності-об'єкти отримують статус невизначених, проте потенційно складених, етимологію яких буде перевірено на наступних кроках.

На наступному кроці в групі, де відібрані атомарні та слабкі сутності-об'єкти, повторно та більш доказово автоматизовано відокремлюють атомарні від слабких завдяки двом критеріям, що одночасно використовують. Перший критерій полягає в тому, що для ідентифікації будь якого значення природного атрибута атомарної сутності-об'єкта достатньо лише назви сутності-об'єкта та назви атрибута, що неможливо саме у випадку слабкої сутності. Але таке співставлення на цьому кроці здійснюється на значно більшій кількості даних. Другий критерій способу має суто математичне походження і полягає в том, що між атрибутами нащадка та сукупними атрибутами всіх предків спостерігається функціональна залежність, а отже детермінований зв'язок, що дає можливість відстежувати не лише сам факт слабкості, а ще й конкретизувати ланки зв'язків з більш старшими сутностями-об'єктами. Причому, якщо зв'язок від нащадка до предка встановлюється однозначно, перевірка наявності або відсутності однозначного зворотного зв'язку від предка до множини нащадків можлива лише завдяки інтерполяції значень від атрибутів всіх нащадків наступного рівня. Тобто перетворення множини цих значень у математичну функцію та перевірку детермінованої залежності на відрізку в околі значень атрибута конкретного нащадка. На зразок від стеження детермінованого зв'язку, наприклад, у періодичній функції. А сама схема інтерполяції - це широковідомі алгоритми, які можна підібрати виходячи зі специфіки предметної області. В переважній більшості випадків достатньо використати певний тип поліноміальної інтерполяції, де аргументами поліномів можуть бути або явний вигляд значень атрибутів, або змінні логічного типу. Підтверджений зв'язок відображається в структурі коміркових ідентифікаторів сутностей-об'єктів.

Проте, якщо на цьому кроці з'ясовується, що деякі сутності-об'єкти, що віднесені до категорії слабких, віднесені помилково, уточнена етимологія кожної потенційно слабкої сутності-об'єкта визначиться на наступному кроці способом, адже така хибність може виникнути лише за причин, що етимології слабких та складених сутностей-об'єктів подібні, що у випадку повільної

5 залежності складеної сутності від часу може привести до помилкової сепарації цієї сутності-об'єкта. А варіант, що атомарна сутність-об'єкт суттєво залежить від часу і тому помилково потрапила до групи складених сутностей-об'єктів, є майже неймовірним. Тому ця ситуація також однозначно визначиться на наступному кроці.

10 Для подальшого уточнення не лише характеру та приналежності до групи складених сутностей, а ще й остаточного відновлення конкретної структури та походження кожної ланки етимології кожної складеної сутності-об'єкта, коли використання способів порівняння за попередніми пунктами не є достатнім, на базі сукупності атомарних та слабких сутностей-об'єктів, отриманої від попередніх кроків способом, в пам'яті в якості шаблону будують каркас повної сукупності зв'язків даних. І в межах цієї синтезованої повної сукупності продовжують

15 подальші ітерації процедури послідовних наближень порівняння потенційних складених сутностей-об'єктів з шаблонними за наступною схемою.

1. На базі груп атомарних та слабких сутностей-об'єктів формують базову сукупність сутностей-об'єктів: до відібраної групи атомарних сутностей-об'єктів додають ще й підгрупу віртуально атомарних сутностей-об'єктів, які одержують додаванням до ідентифікаторів слабких

20 сутностей-об'єктів окремого унарного ідентифікатора, немов би вона - атомарна, створюючи в такий спосіб початкову множину простих унарних ідентифікаторів. Ця дія носить суто технологічний характер і спрощує подальші дії щодо створення комбінацій коміркових ідентифікаторів. Адже призначені віртуально атомарні сутності-об'єкти, що походять від слабких, несуть в собі обидві етимології - природну, тобто складену, та штучну, тобто унарну.

25 Але це не призводить до суперечностей ні при маніпулюванні даними, ні при відстеженні цілісності даних, ані при подальших модифікаціях, оскільки в кожній віртуальній сутності-об'єкті зберігається детермінований бінарний зв'язок між природним складеним комірковим ідентифікатором та штучним унарним. Цей же зв'язок можна відстежити і в усіх наслідкових складених сутностях-об'єктах, що синтезуються на подальших кроках способу. Це докорінно

30 відрізняє таку процедуру в способі, що заявляється, від процедури автоматичного призначення унарного ідентифікатора без врахувань семантики будь-якому об'єктові, що притаманно, наприклад, об'єктно-орієнтованій моделі.

2. Для кожного унарного ідентифікатора кожної сутності-об'єкта з базової сукупності відводять у сховище одинарний домен пам'яті для розміщення елементів зберігання ідентифікатора, структура якого строго унарна. В такий спосіб в пам'яті створюється початкова

35 множина простих одинарних доменів. При цьому ідентифікатори від слабких сутностей-об'єктів можуть бути позначеними додатково. Проте спосіб установки подібних міток може бути будь-яким, навіть до їх відсутності.

3. У сховищі синтезують каркас-шаблон еталонних складених сутностей-об'єктів, для чого здійснюють сполучення Декартових перемножень згаданих одинарних ідентифікаторів між собою за принципом "всі на всі". Цією процедурою породжують систему доменів з багатоарними ідентифікаторами, структура кожного з яких строго відповідає структурі функціональної частини відповідних синтезованих складених предикатів. При цьому структура деяких з них відповідає

40 структурі слабких сутностей-об'єктів з третьої групи способу. Після цього синхронізовано заповнюють відповідними даними семантично сумісні домени. Цим одержують повну сукупність семантичних сполучень складених доменів, що означає, що у цій синтезованій сукупності кожні К-арні складені домени породжені декартовим добутком К примірників атомарних (або віртуально атомарних, тобто слабких сутностей-об'єктів - на цьому кроці це не має суттєвого значення) сутностей-об'єктів, тобто К-ю вибіркою з базової сукупності. Це і синтезує повний

45 каркас поіменованих структурованих комірок під розміщення даних від атрибутів складених сутностей-об'єктів з початкового потоку. Саме тому такий каркас може бути використаним як шаблон. При цьому загальне число таких складених доменів з ідентифікаційними комірками, а в подальшому і таблиць із даними, дорівнює кількості множин булеану, тобто сполучень множин всіх підмножин. На цьому кроці в комірки синтезованого каркасу-шаблону розміщують значення

50 всіх одержаних з початкового потоку опису предметної області атрибутів з урахуванням віднайдених етимологій, тобто, коміркових ідентифікаторів.

4. Завдяки процедурам статистичного аналізу з використанням конкретних значень даних здійснюють остаточну перевірку груп атрибутів атомарних, складених та слабких сутностей-об'єктів з початкового потоку та сформованих атомарних та складених ідентифікаторів на

60 відповідність один одному. При цьому способом передбачається можливість багаторазового

уточнення цієї відповідності шляхом застосування повторної процедури послідовних наближень та багаторазової модифікації базової сукупності та відповідного каркасу-шаблону, що в остаточному підсумку приведе до повного збігу етимології всіх сутностей-об'єктів з початкового потоку з етимологією штучно синтезованих на каркасі.

5 Спосіб передбачає можливість розвитку процедур логічного та статистичного аналізу. Для цього окремо будується зовнішня бібліотека, яка поповнюється новими підпорядкованими способами як логічного, так і статистичного аналізу зі своїми новими критеріями, що розробляються користувачами. Тому перелік підпорядкованих способів порівняння даних між собою, а також перелік критеріїв порівняння нічим не обмежується. Не обмежується і

10 послідовність виконання згаданих процедур. Проте, якщо словник імовірних етимологій на початкових стадіях свого існування не є повним, постійна експлуатація, що поповнює його, мінімізує потребу в автоматизованому логічному та статистичному аналізі вхідних потоків.

В теорії каркасної моделі доводяться теореми повноти та несуперечливого розвитку каркасу, побудованого на булеані базової сукупності сутностей-об'єктів, основним наслідком яких є висновок, що складені сутності-об'єкти подальших зв'язків не утворюють та наступних сутностей-об'єктів не породжують. Не складно довести, що, якщо будь-якій сукупності складених сутностей-об'єктів штучно призначити статус атомарних зі штучними унарними ідентифікаторами, то знов утворені нові (штучні) складені сутності-об'єкти (тобто, по суті, - їх зв'язки) можна отримати і на "попередньому" каркасі, тобто без цього перепозначення

20 ідентифікаторів. Тобто, базова сукупність сутностей-об'єктів - це ще й базова сукупність ідентифікаторів. А як зрозуміло з викладеного, синтезовані складені сутності-об'єкти не розширюють базової сукупності. Проте будь-яке розширення базової сукупності призводить до появи нових складених сутностей-об'єктів. Тому користувач, якщо виникає потреба, має технологічну можливість штучно моделювати такі подальші зв'язки саме розширенням базової

25 сукупності ідентифікаторів - додаючи до базової сукупності ще й штучні ідентифікатори.

Для побудови сховищ даних, які б мали властивість швидкісного виконання як реляційних, так і об'єктно-орієнтованих запитів, колена атомарна ознака кожної сутності-об'єкта, тобто кожна атомарна сукупність даних, що об'єднується одномісною частиною багатомісного у загальному випадку предиката в атрибут цієї сутності-об'єкта, наділяється власним

30 структурованим унікальним ідентифікатором. Спільна частина структури цього ідентифікатора будується у відповідності до структури етимології сутності-об'єкта, тобто структури функціональної частини багатомісного предиката. А остання, унікальна ланка ідентифікатора відповідає значенням даних цієї атомарної ознаки. Це дає можливість виконувати запити з використанням індексації ідентифікатора за його структурою, що значно збільшує швидкість

35 одержання відповіді. Та, в свою чергу, дає можливість об'єднати властивості табличної та нетабличної форм зберігання. Цю нетипову форму отримують завдяки нетабличному об'єднанню сукупностей даних в атрибути сутностей-об'єктів у відповідності до спільних за найменуванням та структурою ідентифікаторів. Ця нова властивість важлива для еволюціонування схеми даних при експлуатації сховища. А також для відповіді на запити від

40 реляційних та не реляційних систем маніпулювання даними, що засновані, наприклад, на об'єктно-орієнтованих моделях даних.

У сховища, побудованого у відповідності до способу, що заявляється, виникає ще одна значна перевага. Надається можливість або відокремленого та паралельного опрацювання кожного даного незалежно одне від одного, або групового опрацювання декількох об'єднаних

45 груп даних, як залежно, так і незалежно одна від одної. При цьому не існує потреби в строгій відповідності кожного даного зі спільного атрибуту один до одного за значенням, тилом та розміром, як того вимагає, наприклад, реляційний спосіб розміщення, оскільки від кожного даного вимагається лише наявність спільного ідентифікатора зі структурою, відповідною структурі спільного предиката.

Запропонованим способом створюється універсальна технологія сепарації даних, яка не залежить від особливостей довільної предметної області та дозволяє в автоматизованому режимі провести аналіз та декомпозицію предметної області будь-якої складності. А це в свою чергу дозволяє: здійснювати автоматизоване каркасне розміщення даних, без переробки експлуатуючого програмного забезпечення мінімально необхідними операціями виконувати

50 будь-які семантично доцільні модифікації структури сховища та формувати набір єдиних процедур обробки даних - групових функцій. І таким чином стандартизувати технологію генерації та експлуатації сховищ даних.

Узагальнену послідовність кроків способу ілюструє блок-схема на Фіг. 1. Суть важливої властивості способу за пунктом No 15 "формули" пояснюється кресленням, де на Фіг. 2 наведено частково заповнену таблицю з рознесеними даними випадковим чином. Тут терміном

60

nilл позначено відсутність даних. Заповнену ж комірку позначено літерою А від слова "атрибут" та індексом, де перша однорозрядна цифра означає номер рядка, а друга дворозрядна - номер стовпчика. Отже, на Фіг. 2 наведено канонічну табличну форму, в якій не зважаючи на порожні комірки, відстежуються всі стовпчики та рядки. На Фіг. 3 наведено оптимізовану форму, де взагалі відсутні порожні комірки, а також подібні за структурою предикатів атрибути не обов'язково повинні мати однакові розміри. Це надає можливість поєднати властивість реляційного та об'єктно-орієнтованого способів розміщення.

В матеріалах заявки використовуються наступні поняття та назви (всі терміни упорядковані не за абеткою, а згідно з логікою використання):

- предикат (одне з можливих значень, що використовується саме в цій заявці) - спільна логічна ознака всіх елементів множини, насамперед множини атрибутів, що надає можливість розрізнити атрибути, а також з'ясувати, до якої сутності-об'єкта цей атрибут належить. Матеріали заявки побудовано на підґрунті каркасної моделі даних, в якій кожний атрибут може мати лише один унікальний предикат, що пов'язує його лише з однією сутністю-об'єктом. В загальному значенні предикат - це функція, яка має лише два логічних значення - "істина" або "хибність" ("так-ні", "свій - чужий" й ті.). В цій моделі предикат може бути складеною функцією, що має багатомісну аргументну та складену функціональну частини. Склад предикату - це кон'юнкція (логічний добуток) простих багатомісних предикатів, одночасне виконання умов кожного з яких повертає спільну "істину", а невиконання умов хоча б одного з них поверне "хибність". Предикат сутності-об'єкта є наслідком та носієм її походження. Розглядається лише два шляхи утворення будь-якої сутності-об'єкта - або породженням атомарними слабких сутностей-об'єктів за принципом "одна породжує багатьох", або внаслідок рівноправних зв'язків між атомарними чи слабкими сутностями-об'єктами, за принципом "багато до багатьох". Проста або складена функціональна частина предиката - це і є наслідок етимології змісту сутності-об'єкта.

- сутність-об'єкт - символ певного атомарного змісту, що кодується словом, тобто по суті, це і є предикат, що поєднує в одну групу множини атрибутів - властивостей сутності-об'єкта. В цій моделі кожна сутність-об'єкт може мати єдиний унікальний природний предикат та декілька штучних;

- довільна предметна область (предметна область довільного обсягу та довільної структури)
- довільна сукупність сутностей-об'єктів, що користувачем сприймається як єдина система, функціонування якої ним досліджується та моделюється;

- атрибут - властивість сутності-об'єкта, що має той самий предикат, що й всі атрибути цієї сутності-об'єкта;

- природний атрибут - властивість, що не надано користувачем предметної області, а виявлено серед сукупності атрибутів сутності-об'єкта завдяки аналізу предметної області;

- штучний атрибут - атрибут, який в структуру сутності-об'єкта штучно введено користувачем предметної області;

- етимологія - походження змісту сутності-об'єкта, що відображається в структурі функціональної частини утворюючого предиката та виражається відповідним сумарним рядком символів. Цей рядок утворює ідентифікатор. При цьому, якщо в граматиці мови у іменника "етимологія" не існує множини, то в логіко-математичному сенсі сутність-об'єкт може мати декілька етимологій. Тому в заявці цей термін використовується ще й в формі, яка не передбачена граматику мови;

- атомарна сутність-об'єкт - сутність-об'єкт, що має унарну етимологію, тобто така, що формується предикатом, який має виключно унарну функціональну частину;

- слабка сутність-об'єкт - сутність-об'єкт, що має складену етимологію, тобто така, що формується предикатом, який має виключно багатоарну функціональну частину, виключаючи унарну, та ще й функціональну, тобто ієрархічну, залежність кожної наступної ланки функціональної частини предиката, виключаючи найстаршу, від сукупності попередніх, тобто від сукупності предикатів-предків;

- складена сутність-об'єкт - сутність-об'єкт, що має складену етимологію, тобто така, що формується предикатом, який має виключно багатоарну функціональну частину, виключаючи унарну. Причому цей предикат не має ніяких функціональних, тобто ієрархічних, залежностей будь-якої ланки функціональної частини одна від одної. Проте спостерігається функціональна залежність сумарної сукупності ланок функціональної частини від сумарної сукупності ланок функціональних частин утворюючих предикатів;

- артефакт - сутність-копія, атрибути якої є копіями атрибутів інших сутностей-об'єктів, причому об'єднання цих атрибутів в цю сутність-об'єкт є штучним, отже кожному з цих атрибутів

штучно призначені додаткові предикати, що забезпечують об'єднання атрибутів у цю штучну сутність-об'єкт;

5 - невизначена сутність-об'єкт - сутність-об'єкт, етимологія якої підлягає подальшому уточненню завдяки додатковій інформації з предметної області, причому до цієї ж групи відбираються і такі сутності-об'єкти, що не мають жодного примірника, а лише в межах певної предметної області мають абстрактне ім'я або поняття, а тому не можуть бути самостійно використані;

- невизначені відокремлені атрибути - поодинокі атрибути, що за рахунок однакового написання іменників у вхідному потоці помилково замасковані під сутності-об'єкти;

10 - базова сукупність сутностей-об'єктів - сукупність лише атомарних та слабких сутностей-об'єктів, причому така, що серед ланок слабких сутностей не існує порожнин, а також така, що для кожної ланки слабких сутностей-об'єктів визначені початкові атомарні предки;

15 - структурований комірковий ідентифікатор - ідентифікатор комірки пам'яті, яка містить дані від того чи іншого атрибута сутності-об'єкта, що має певну типізовану структуру. В межах способу ця структура строго відповідає структурі етимології сутності - об'єкта і, отже, етимології атрибута. Тому її призначає не користувач, а автоматизовано призначає окрема процедура способу під час виконання сепарації, адже саме цей ідентифікатор і є наслідком шуканої сепарації;

20 - рядкове об'єднання (рядкова сума) - одержання нового ідентифікатора з ідентифікаторів-частин завдяки їх лінійному об'єднанню на зразок утворення слів завдяки рядковій сумі букв. При цьому в деяких випадках місцезнаходження літер в ідентифікаторі не має значення, як наприклад в ідентифікаторі атрибутів складених сутностей-об'єктів. А у випадках слабких сутностей-об'єктів місцезнаходження частини ідентифікатора означає напрямок залежності. Як правило, це кодується зліва на право, тобто крайня ліва частина символізує початкову атомарну сутність-об'єкт. Наприклад, рядкова сума літер "с", "п", "о", "с", "і" та "б" поверне слово "спосіб", якщо це - слабка сутність-об'єкт;

25 - слово (іменник та дієслово) - унікальний набір літер, що одночасно використовується і як унікальна назва сутності-об'єкта або зв'язку в пам'яті, і як їх назва в мовному описі предметної області, з якою працює користувач.

30 Допоміжні слова, без яких речення може не мати мовного змісту, належать до дієслів та обумовлюють клас зв'язку;

- речення (атомарне речення) - зв'язок (бінарний) між двома сутностями-об'єктами, причому складні речення, тобто речення, що описують декілька бінарних або навіть багатоарні зв'язки мають бути декомпозовані на декілька атомарних;

35 - початковий потік опису предметної області - повна сукупність атомарних речень, які описують предметну область з урахуванням всіх початкових файлів - звукових, текстових, фалів схем та навіть файлів сховищ даних, що вже існують та впроваджені;

40 - автоматизований логічний аналіз - процедура логічного співставлення найменувань сутностей-об'єктів зі словником імовірних етимологій, а також врахування всіх наданих у початковому потоці зв'язків їх між собою, без використання безпосередніх значень атрибутів та без використання математичних критеріїв виявлення детермінованих залежностей множин даних та математичної близькості даних між собою;

45 - автоматизований статистичний аналіз - процедура математичного співставлення значень атрибутів сутностей-об'єктів між собою з використанням математичних критеріїв на предмет виявлення детермінованих залежностей між множинами даних від атрибутів та виявлення математичної близькості даних між собою;

- булеан - термін з формальної логіки, що означає множину всіх підмножин, тобто повне комбінаторне сполучення множини будь яких елементів;

50 ФОРМУЛА ВИНАХОДУ

1. Спосіб попередньої каркасної сепарації даних перед їх модифікаційно здатним розміщенням у сховище або процесом подальшої обробки, який полягає в тому, що розміщені дані розподіляють на групи згідно з семантикою змісту сутностей або об'єктів, які об'єднують кожну таку групу, і така група даних має спільний набір характеристик, що відповідають спільному предикату, а групи сутностей-об'єктів перебувають між собою у різноманітних відносинах, коли для вхідного аналізу даних використовуються онтології, тобто словники абстрактних предметних областей, побудовані за тими чи іншими факторами, причому враховуються всі можливі часткові копії кожної сутності-об'єкта, утворюючи маски цих сутностей-об'єктів, після чого моделюються всі зв'язки між групами цих масок сутностей-об'єктів у предметній області,

для чого кожній групі масок відводять у сховищі кілька ділянок пам'яті для розміщення елементів зберігання, створюючи, таким чином, розширену початкову множину ділянок пам'яті, і схема сховища будується в цифровій пам'яті відповідно до сполучень декартових добутків всіх доменів-масок між собою за принципом "всі на всі", формуючи при цьому каркас зв'язків доменів-масок, при цьому загальна кількість розташовуваних груп атрибутів доменів-масок, істотно збільшується і відповідає множині всіх підмножин зв'язків доменів-масок сутностей-об'єктів, який **відрізняється** тим, що отриманий сполученнями декартових добутків сутностей-об'єктів каркас зв'язків приймають за нульове наближення як шаблон, на якому для врахування семантики довільної предметної області здійснюють автоматизований логічний і статистичний аналіз опису довільної предметної області, а подальші послідовні наближення автоматизовано розміщують дані в сховище й автоматизовано виключають семантично несумісні групи атрибутів.

2. Спосіб за п. 1, який **відрізняється** тим, що для семантичного аналізу опису довільної предметної області здійснюють зчитування:

або звукового голосового сигналу в реальному часі,

або файла з записаним голосовим сигналом, що надиктовані природною мовою та описують довільну предметну область,

або текстового файла опису предметної області, сформованого текстом природною мовою,

або файла, сформованого мовою послідовних схем або графів, що відповідають опису предметної області,

або послідовності файлів сховищ даних, що вже існують та впроваджені в експлуатацію,

а також перетворення в словесний потік та розпізнавання всіх відокремлених слів у початковому потоці, при цьому для подальшого автоматизованого аналізу кожний цифровий потік використовується не лише відокремлено, а й у порівнянні один з іншим, після цього здійснюють розміщення всіх отриманих слів у пам'яті.

3. Спосіб за п. 2, який **відрізняється** тим, що по чергово автоматизовано аналізується кожне слово за принципом послідовних наближень з можливістю динамічного врахування додаткових відомостей про дані із предметної області, причому сукупний початковий потік у пам'яті перетворюють в потік, що має наступну форму: технологічною одиницею початкового потоку для автоматизованого аналізу є одне атомарне речення, кожне з сукупності яких містить тільки дві сутності-об'єкти, кожна з яких кодується іменником з унікальним політерним написанням, причому так, що іменники, які повторюються, означають ту саму сутність-об'єкт, тому таке повторення в межах одного речення означає тривіальну пару, тобто, таку, котра несе лише декларацію про наявність цієї сутності-об'єкта, без зв'язків її з іншими, і дієсловом між ними, що означає бінарний зв'язок між парою сутностей-об'єктів, з унікальним політерним написанням так, що дієслова, які повторюються, означають той самий клас зв'язку, при цьому кількість речень не обмежується змістом предметної області, також здійснюють попередній формальний аналіз наявності для кожної задекларованої сутності-об'єкта щонайменше одного зв'язку з будь-якою іншою сутністю-об'єктом.

4. Спосіб за п. 3, який **відрізняється** тим, що для перетворення файла початкового потоку опису предметної області, сформованого мовою послідовних схем або графів, у потік слів, кожній графовій фігурі схеми ставиться у відповідність іменник, а дузі графа ставиться у відповідність дієслово, причому виконують окрему процедуру строгого відокремлення зі схемного початкового потоку пар сутностей-об'єктів та їхніх зв'язків, а також визначення їх іменниками й дієсловами, тобто опрацювання графових схем типу ER-схем з урахуванням обмежень унікальності політерного найменування сутностей-об'єктів, такі ж самі дії виконують і при перетворенні в атомарні речення файлів сховищ даних.

5. Спосіб за п. 1, який **відрізняється** тим, що в пам'яті формують окрему ділянку, де розміщуються структуровані коміркові ідентифікатори, структура кожного з яких строго відповідає імовірній семантичній структурі змісту кожної сутності-об'єкта, яка автоматизовано відстежується завдяки критеріям, які побудовані на єдиному узагальненому факторі - походженні змісту сутності-об'єкта, тобто її етимології, причому в будь-якій предметній області будь-якого обсягу та структури всі сутності-об'єкти розподіляють на три категорії - атомарні сутності-об'єкти, слабкі та складені після зв'язкові сутності-об'єкти, а породження сутностей-об'єктів здійснюється за наступною схемою: на базі атомарних породжуються слабкі, тобто функціонально залежні від атомарних, причому така залежність може бути як лише на рівні ідентифікації слабких атрибутів, так і на рівні всього існування залежних слабких сутностей-об'єктів, на базі ж сукупної групи атомарних та слабких сутностей-об'єктів завдяки утворенню зв'язків між ними створюються складені після зв'язкові сутності-об'єкти, причому процес утворення слабких та складених сутностей-об'єктів маскують частини мови - іменники та

терміни, що їм відповідають, всі інші фактори, що характеризують семантику змісту будь-якої сутності-об'єкта в довільній предметній області є функціонально залежними від етимології, яка, в свою чергу, описується математичною логікою предикатів та у вигляді рядкового структурованого коміркового ідентифікатора має наступну загальну схему:

$$5 \quad X_1^{m_1} + X_2^{m_2} + X_3^{m_3} + \dots + X_{k_j}^{m_k},$$

де кожна ланка $X_{k_j}^{m_k}$ - відокремлений ідентифікатор факту походження j -ї сутності-об'єкта, k_j - номер ланки ідентифікатора j -ї сутності-об'єкта, m_k - номер відповідної породжуючої сутності-об'єкта з об'єднаної групи атомарних та слабких сутностей-об'єктів, причому кожне m_k може отримати будь-яке значення лише з множини $\{1, 2, \dots, N_0, \dots, N\}$, де N_0 - загальна кількість

10 атомарних сутностей-об'єктів, N - сумарна кількість атомарних та слабких сутностей, j - номер довільної сутності-об'єкта в довільній предметній області, причому у випадку повної сукупності зв'язків $j = \{1, 2, \dots, N, \dots, (2^N - 1)\}$, де знак "плюс" означає рядкове об'єднання, причому для атомарних сутностей етимологією є лише одна ланка X^j , в якій $m = j$, тобто атомарна сутність породжує сама себе, де атомарні сутності-об'єкти одержують в загальній сукупності перші

15 номери, тобто для них $j = 1, N_0$, для слабких сутностей етимологією є вищезгадана рядкова сума ланок, де кожному номеру k_j ланка $X_{k_j}^{m_k}$ відповідає строго, тобто послідовність ланок строго відповідає послідовності залежностей кожної наступної ланки від попередньої, що в свою чергу відповідає послідовності синтезу кожною попередньою слабкою сутністю-об'єктом, аж до найстаршої атомарної, наступної слабкої сутності-об'єкта, для складених сутностей-об'єктів

20 етимологією є також рядкова сума ланок, де місце кожної ланки $X_{k_j}^{m_k}$ не строге, тобто послідовність ланок не має значення, проте сумарна сукупність ланок строго відповідає сукупності утворюючих сутностей-об'єктів, тож в загальному випадку для будь-якої сутності-об'єкта весь структурований комірковий ідентифікатор являє собою сумарний рядок літер або цифр, кожна ланка яких має мінімально достатній рядковий розмір, причому структурований

25 комірковий ідентифікатор унікально ідентифікує всі властивості конкретної сутності-об'єкта, тобто її атрибути.

6. Спосіб за п. 5, який **відрізняється** тим, що дані піддають початковій фазі автоматизованого логічного аналізу, тобто, початковий потік слів підготовчими автоматизованими процедурами розподіляється на групи:

30 атомарних сутностей-об'єктів з унарною етимологією, слабких сутностей-об'єктів, які мають складену етимологію, складених сутностей-об'єктів, які мають складену етимологію, артефактів, тобто сутностей-копій, дані від яких копіюють дані від атрибутів інших сутностей-об'єктів,

35 невизначених сутностей-об'єктів або окремих атрибутів, семантика яких підлягає подальшому уточненню завдяки додатковій інформації із предметної області, причому в подальшому в ідентифіковані комірки сховища розміщують групи атрибутів сутностей-об'єктів, причому унарні ідентифікатори комірок сховища строго відповідають атомарним сутностям-об'єктам, а складені ідентифікатори комірок строго відповідають слабким та складеним сутностям-об'єктам.

40 7. Спосіб за п. 6, який **відрізняється** тим, що в пам'яті здійснюють послідовне або одночасне, тобто паралельне, виконання для кожної сутності-об'єкта з кожного речення, тобто, з кожної пари, процедури порівняння з кожною іншою сутністю-об'єктом, причому ця процедура порівняння виконує відокремлені підпорядковані способи автоматизованого логічного

45 вичленовування замаскованої етимології кожної сутності-об'єкта, тобто семантичної структури її змісту, результатом виконання яких є шукана сепарація, тобто, надання кожній комірці, де зберігаються дані від атрибутів кожної сутності-об'єкта з початкового потоку відповідних структурованих коміркових ідентифікаторів та перегрупування сутностей-об'єктів в сховищі у окремо розміщені групи, при цьому відновлення етимології кожної ланки сутностей-об'єктів

50 здійснюють автоматизованим логічним аналізом іменників та дієслів, тобто аналізом змісту сутностей-об'єктів та змісту зв'язків, без врахування множин конкретних значень конкретних атрибутів сутностей-об'єктів, причому аналіз засновано на порівнянні змісту сутностей-об'єктів між собою за принципом "всі з усіма" з використанням словника імовірних етимологій змісту сутностей-об'єктів, де кожному іменнику заздалегідь поставлено у відповідність найімовірнішу

етимологію, тобто встановлюється відповідність між словами з початкового потоку та словами, що існують в словнику, а результатом такого співставлення є перше наближення шуканої сепарації сутностей-об'єктів, а також отримання першого наближення структур їх етимології.

8. Спосіб за п. 7, який **відрізняється** тим, що слова, які позначають невідомі для словника сутності-об'єкти та класи зв'язків, для подальшого автоматизованого аналізу відокремлюються в пам'яті, причому завдяки різним критеріям відстежується етимологія невідомих словникові сутностей-об'єктів, та пропонуються рекомендації щодо знайдених логічних помилок та протиріч у початковому потоці щодо можливих некоректних використань іменників та дієслів, що означають нелогічності в роботі певних ділянок предметних областей.

9. Спосіб за п. 8, який **відрізняється** тим, що проводять автоматизований логічний аналіз невідомих словнику імовірних етимологій сутностей-об'єктів, причому спочатку відокремлюються невідомі потенційні складені сутності-об'єкти завдяки автоматизованому логічному порівнянню кожної з невідомих сутностей-об'єктів з тими, що утворюються з повторюваних іменників і повторюваних дієслів з початкового потоку завдяки об'єднанню їх в одну складену, тобто багатобічну післязв'язкову сутність-об'єкт, за умови збігу класу зв'язку, тобто збігу дієслів між різними парами, причому наявність невизначених сутностей-об'єктів, що мають логічні суперечності, та артефактів в цих попередньо відібраних групах сутностей-об'єктів ігнорується.

10. Спосіб за п. 9, який **відрізняється** тим, що здійснюють завершальну фазу автоматизованого логічного аналізу початкового потоку, для чого автоматизовано аналізуються групи тих сутностей-об'єктів та зв'язків, що виявилися невідомими словнику імовірних етимологій та залишилися після вилучення потенційно складених сутностей-об'єктів, причому невідомі атомарні сутності-об'єкти відокремлюються з використанням єдиного логічного критерію, який полягає в тому, що в загальному випадку для ідентифікації будь-якого значення природного атрибута атомарної сутності-об'єкта достатньо лише назви сутності-об'єкта та назви цього атрибута та додається інформація щодо природних атрибутів кожної з сутностей-об'єктів, що аналізуються, а також значення кожного з цих атрибутів, причому кожна сутність-об'єкт отримує статус або атомарної сутності-об'єкта, або слабкої, або невизначеної, причому наявність артефактів ігнорується і вони також одержують один зі згаданих статусів.

11. Спосіб за п. 10, який **відрізняється** тим, що, якщо після автоматизованого логічного аналізу початкового потоку сутностей-об'єктів та зв'язків група невизначених сутностей-об'єктів не стає порожньою, тобто завдяки автоматизованому логічному аналізу неможливо віднести ці сутності-об'єкти до згаданих трьох категорій, кожній з цих суперечливих сутностей-об'єктів примусово призначається статус атомарної, але на рівні їхнього коміркового ідентифікатора обов'язково позначається це, додаючи до унарного ідентифікатора спеціалізовану відокремлену ланку, відповідальну за цю особливість, і формуючи тим самим окрему підгрупу суперечливих сутностей-об'єктів в групі атомарних, що при подальшій експлуатації сховища при потребі модифікації його структури дає можливість внести відповідні корегування.

12. Спосіб за п. 11, який **відрізняється** тим, що серед попередньо відібраних груп сутностей-об'єктів остаточно відокремлюються артефакти, тобто сутності-копії, для чого здійснюють автоматизоване статистичне порівняння, засноване на використанні процедур статистичного аналізу для виявлення детермінованих функціональних або кореляційних чи регресивних багатозначних залежностей між значеннями даних в атрибутах сутностей-об'єктів, тобто при виникненні прямих збігів назв груп атрибутів, а також і їх значень у різних сутностей-об'єктів, на рівні їх коміркових ідентифікаторів окремо фіксують та визначаються щодо зберігання надлишкових даних, при цьому, коли назви атрибутів, що належать різним сутностям-об'єктам, є різними, а їх значення ідентичні, що з'ясовується на збільшеній кількості значень атрибутів, це відображається в структурі коміркового ідентифікатора.

13. Спосіб за п. 12, який **відрізняється** тим, що будують уточнене наближення сепарації складених сутностей-об'єктів, для чого враховується, що для коректності статистичного аналізу вся сукупність значень всіх атрибутів від всіх сутностей-об'єктів предметної області мають відповідати єдиному моменту часу життя предметної області, причому відстань між сусідніми проміжками часу має бути достатньою для виникнення дійсно нового стану предметної області, для чого відокремлюють групи значень атрибутів, що залежать від часу, та групи значень атрибутів, що від часу не залежать, або якщо і залежать, то лише від дуже значних проміжків часу - розвитком та змінами яких можна знехтувати у порівнянні з іншими групами значень атрибутів, причому групу атрибутів, яка не залежить від часу, відокремлюють до групи сутностей-об'єктів, що створюють структуру предметної області, причому за уточнене чергове наближення складених сутностей-об'єктів береться група сутностей-об'єктів, які залежать від часу, а інша група отримує статус сукупності атомарних, атомарних-невизначених та слабких,

після чого кожна складена сутність-об'єкт зі щойно отриманої групи зіставляється з групою складених сутностей-об'єктів, що залишилися після автоматизованого логічного аналізу, причому в процедурі порівняння між сумарним значенням кожного примірника загальної сукупності всіх відокремлених атрибутів предків та значеннями примірників будь-якого з атрибутів складених сутностей-об'єктів виникає детермінований функціональний зв'язок, що є достатнім критерієм для ідентифікації та відокремлення складених сутностей-об'єктів, отже, якщо при порівнянні спостерігаються збіги, коміркові ідентифікатори залишаються без змін, в іншому ж випадку у кожній з потенційно складених сутностей, отриманих різними кроками, утворюються два відповідних незалежних коміркових ідентифікатори, що фіксують цю обставину, а ці сутності-об'єкти отримують статус невизначених, проте потенційно складених, що перевіряється на наступних кроках або змушує надати додаткову інформацію.

14. Спосіб за п. 13, який **відрізняється** тим, що в групі, де відібрані атомарні та слабкі сутності-об'єкти повторно автоматизовано відокремлюють атомарні від слабких завдяки двом критеріям, що одночасно використовують: перший критерій полягає в тому, що для ідентифікації будь-якого значення природного атрибута атомарної сутності-об'єкта достатньо лише назви сутності-об'єкта та назви атрибута, проте співставлення здійснюється на збільшеній кількості даних, другий критерій має математичне походження - між атрибутами нащадка та сукупними атрибутами всіх предків є функціональна залежність, а отже детермінований зв'язок, що дає можливість відстежувати не лише сам факт слабкості, а ще й конкретизувати ланки зв'язків з більш старшими сутностями-об'єктами, що відображається в структурі їх коміркових ідентифікаторів, причому, якщо зв'язок від нащадка до предка встановлюється однозначно, перевірка наявності або відсутності однозначного зворотного зв'язку від предка до множини нащадків можлива лише завдяки інтерполяції значень від атрибутів всіх нащадків наступного рівня, тобто перетворення множини цих значень у математичну функцію та перевірку детермінованої залежності на відрізок в околі значень атрибута конкретного нащадка, підтверджений зв'язок відображають в структурі коміркових ідентифікаторів сутностей-об'єктів.

15. Спосіб за п. 14, який **відрізняється** тим, що для подальшого уточнення характеру та приналежності до групи складених сутностей, а також відновлення структури та кожної ланки етимології кожної складеної сутності-об'єкта, на базі сукупності атомарних та слабких сутностей-об'єктів, в пам'яті як шаблон будується каркас повної сукупності зв'язків даних і в межах цієї повної сукупності продовжуються подальші ітерації для порівняння потенційних складених сутностей-об'єктів з шаблонними за наступною схемою:

на базі груп атомарних та слабких сутностей-об'єктів формується базова сукупності сутностей-об'єктів: до відібраної групи атомарних сутностей-об'єктів додається ще й підгрупа віртуально атомарних сутностей-об'єктів, які одержуються додаванням до ідентифікаторів слабких сутностей-об'єктів окремого унарного ідентифікатора, створюючи початкову множину простих унарних ідентифікаторів,

для кожного унарного ідентифікатора кожної сутності-об'єкта з базової сукупності відводиться у сховище одинарний домен пам'яті для розміщення елементів зберігання ідентифікатора, структура якого строго унарна,

в пам'яті створюється початкова множина простих одинарних доменів, при цьому ідентифікатори від слабких сутностей-об'єктів позначаються додатково,

у сховищі синтезується каркас-шаблон еталонних складених сутностей-об'єктів, для чого здійснюється сполучення декартових перемножувальних згаданих одинарних ідентифікаторів між собою за принципом "всі на всі", чим формується система доменів з багатоарними ідентифікаторами, при цьому структура деяких з них відповідає структурі слабких сутностей-об'єктів,

синхронізовано заповнюються відповідними даними семантично сумісні домени, чим одержують повну сукупність семантичних сполучень складених доменів, що означає, що у цій синтезованій сукупності кожні K-арні складені домени породжені декартовим добутком K примірників атомарних сутностей-об'єктів, тобто K-ю вибіркою з базової сукупності,

в комірки синтезованого каркасу-шаблону розміщують значення всіх одержаних з початкового потоку опису предметної області атрибутів з урахуванням віднайдених етимологій, тобто, коміркових ідентифікаторів,

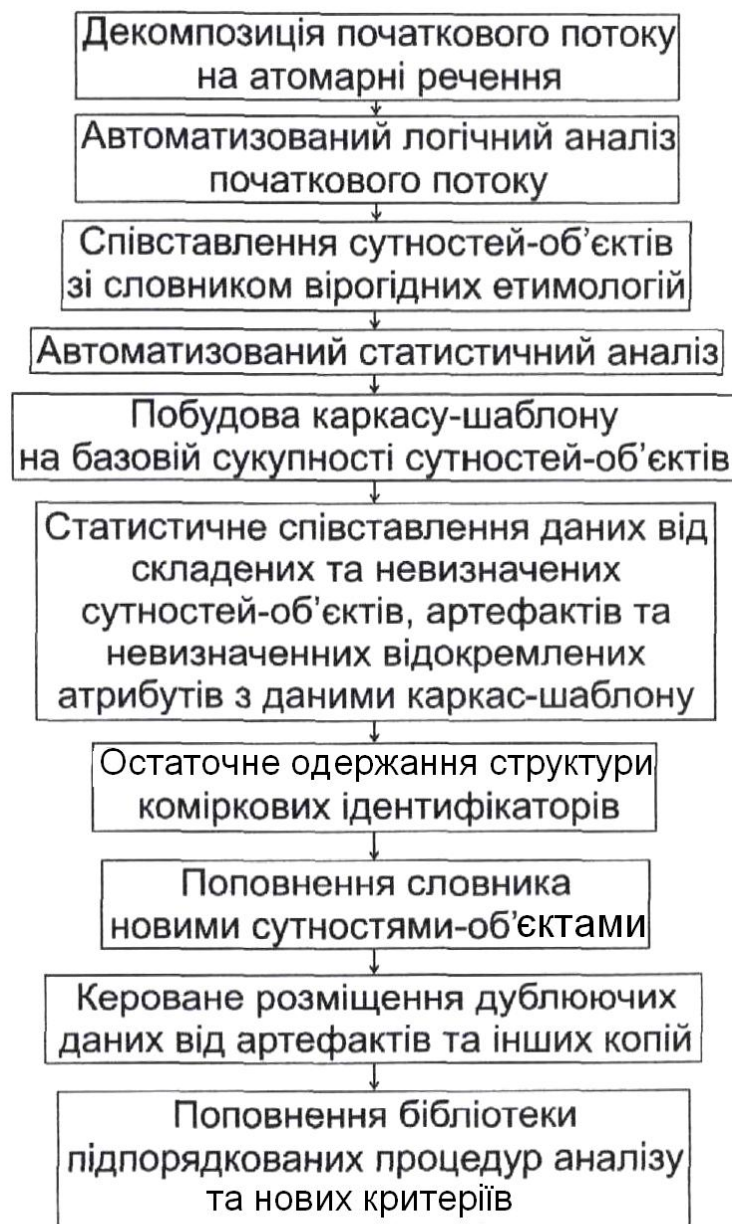
процедурами статистичного аналізу з використанням конкретних значень даних здійснюють остаточну перевірку груп атрибутів атомарних, складених та слабких сутностей-об'єктів з початкового потоку, а також атомарних та складених ідентифікаторів на відповідність один одному,

багаторазово уточнюється відповідність шляхом застосування повторної процедури послідовних наближень та багаторазової модифікації базової сукупності та відповідного

каркасу-шаблону, до повного збігу етимології всіх сутностей-об'єктів з початкового потоку з етимологією штучно синтезованих на каркасі.

16. Спосіб за п. 15, який **відрізняється** тим, що для побудови сховищ даних, які мають властивість швидкісного виконання як реляційних, так і об'єктно-орієнтованих запитів, кожна атомарна ознака кожної сутності-об'єкта, тобто кожна атомарна сукупність даних, має власний унікальний структурований ідентифікатор, спільна частина структури якого є тотожною структурі етимології сутності-об'єкта, а остання, унікальна ланка ідентифікатора відповідає значенням даних цієї атомарної ознаки,

17. Спосіб за п. 16, який **відрізняється** тим, що побудова сховища здійснюється за допомогою об'єднання властивостей табличної та нетабличної форм зберігання, яку отримують завдяки нетабличному об'єднанню сукупностей даних в атрибути сутностей-об'єктів у відповідності до спільних за найменуванням та структурою ідентифікаторів, після чого відокремлено та паралельного індексується кожне дане незалежно одне від одного, або індексується групою, як залежно, так і незалежно одна від одної.



Фіг. 1

A ₁₀₁	A ₁₀₂	A ₁₀₃	A ₁₀₄	Null	Null	Null	Null	Null	Null	Null
A ₂₀₁	A ₂₀₂	Null	A ₂₀₄	Null	Null	A ₂₀₇	Null	A ₂₀₉	Null	A ₂₁₁
Null	A ₃₀₂	Null	Null	Null	Null	A ₃₀₇	Null	A ₃₀₉	Null	Null
Null	Null	Null	Null	Null	Null	Null	A ₄₀₈	A ₄₀₉	Null	Null

...

Fig. 2

A ₁₀₁	A ₁₀₂	A ₁₀₃	A ₁₀₄		
A ₂₀₁	A ₂₀₂	A ₂₀₄	A ₂₀₇	A ₂₀₉	A ₂₁₁
A ₃₀₂	A ₃₀₇	A ₃₀₉			
A ₄₀₈	A ₄₀₉				

...

Fig. 3